COMP3222 Indicative Coursework Marking Scheme (worth 50%)

Updated by Indu Bodala (i.p.boadala@soton.ac.uk) on 20 Oct. 23

Introduction and Data Analysis (7)
(0) There is no introduction or data analysis
(2) The problem and data description is too brief, not very clear or incomplete.
(3) The problem and data are described fully, but explicit characteristics are missing (e.g. data format, data volume, data quality, data bias, feature analysis, computation speed needed, number of classification classes needed).
(5) The problem and data are described fully, with an incomplete list of explicit characteristics.
(7) The problem and data are described fully, with a comprehensive list of explicit characteristics.

Pipeline Design (10)
(0) There is no data pipeline design described (e.g. pre-processing, feature selection, dimensionality reduction and machine learning algorithm).
(3) The data pipeline design and the two algorithms are described, but they are too brief, not very clear, incomplete (both designs must be described) or do not reference the techniques described in the course text and wider literature.
(6) The data pipeline design and the two algorithms are described fully (so the reader could reproduce work if needed), but there are either unreasonable or missing justifications given for the design choices made.
(8) The data pipeline design and the two algorithms are described fully, and reasonable justifications are provided for all design choices made (e.g. explaining why a choice was made in the context of the wider options available and data characteristics).
(10) The data pipeline design and the two algorithms are described fully, reasonable justifications are provided for all design choices made and reference to work in the wider literature is cited for extra context.

Evaluation (10)
(0) There is no evaluation of the pipelines designed and the two algorithms described.
(3) An evaluation is described, but it is too brief, not very clear, incomplete or lacks rigour.
(6) The evaluation is fully described with some rigour, but there are weaknesses in the method used (e.g., metrics used are not defined well enough, only 1 configuration of the algorithm with no exploration).
(8) The evaluation is fully described, and the method is rigorous with multiple algorithm configurations explored, evidence of incremental improvement and statistical significance of the results reported.
(10) The evaluation is fully described, and the method is rigorous with multiple algorithm configurations reported, evidence of incremental improvement and statistical significance of

the results reported. In addition, results are compared to the wider literature cited for extra context.

Conclusion (8)
(0) There is no conclusion
(4) The conclusion summarizes the methods and evaluation results but does not describe any insights gained, or ideas for future improvement.
(8) The conclusion summarizes the method and evaluation results and includes information on insights gained and ideas for future improvement.

Machine learning algorithm implementation (15)
The algorithm implementation will be assessed for clarity, organization and correctness.
Clarity (5 marks) - the code clearly describes the various steps with explanatory comments.
Organization (5 marks) - multiple steps of the pipeline are clearly organized.
Correctness (5 marks) - the code is executable and the F1 scores match the ones reported.

Any detected plagiarism will be reported to the Academic Integrity Officer.